

Data Wrangling with SQL

Good morning!

We start at 10:00

Stock Item Query (1)

Website

item 1
item 2
USB 6 25\$
USB 19\$

17\$

(BAD)

where $sl. UPrice = 25$
 $sl. UPrice$

Python script
(sql query)

Report

list of items
priced more
than USB
Green

(Good)

where $sl. UPrice \geq$ USB Green
Price

NXT StockItem Query (2)

S1

$\geq = ?$

S2

25

Office Cube Pri... B.
DBA joke mng W.
DBA joke mng B.
DBA joke mng dan W.
DBA joke mng dan B.

USB Missile Green 25
USB Rocket Launcher Gray 37
USB Food Flash Drive Sushi 32
USB Food Flash Drive HD
USB Food Flash Drive PS 31

String manipulation (1)

USB Green \longrightarrow USB %
wildcard character

Green USB \longrightarrow % USB

Rocket USB Green \longrightarrow % USB %
Rocket USB blue

String manipulation (2)

⚠ Watch out for the Collation
(sebastien)

CI: "SEBASTIEN" = "sebastien"
= "Sebastien"
= "SEBASTIEN" (sebastien)

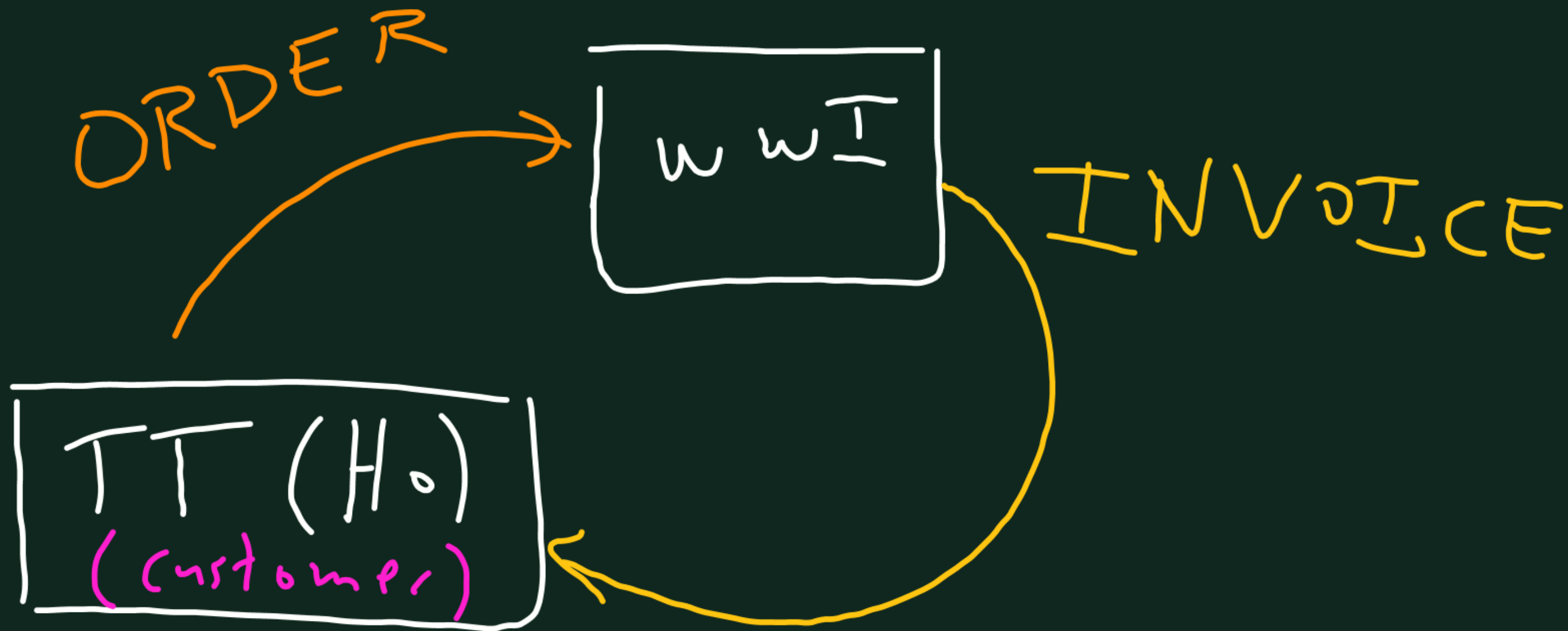
"USB" = "usb"

Like "USB%" = Like "usb%"

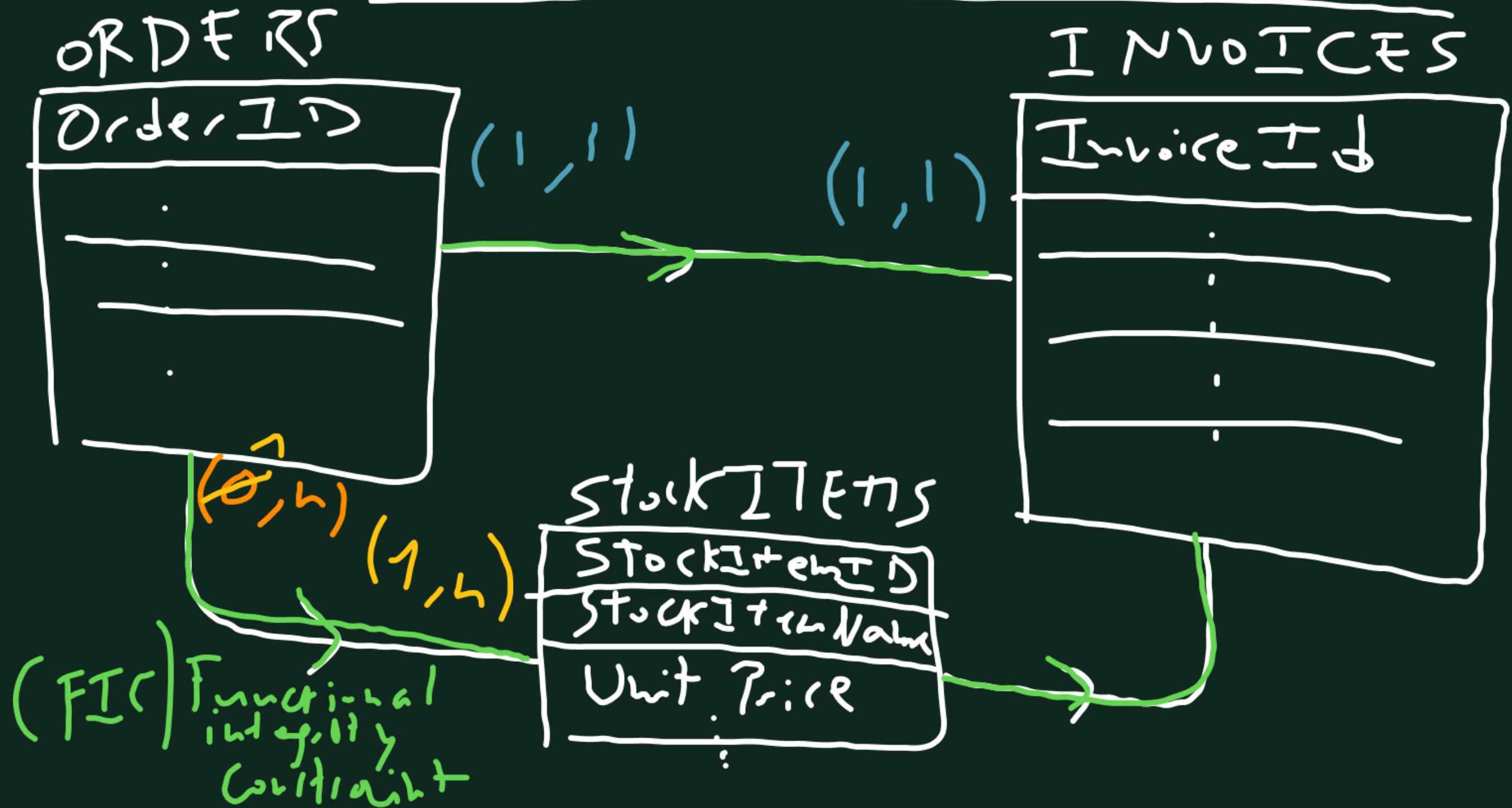
Break

Back at 11:40

Orders & Invoices



Entity/Relationship Diagram



Lunch

Back at 14:00

Orders not invoiced (1)

ORDERS

INVOICES



ORDERS not INVOICED
(Left outer
join)

ORDERS That have
been invoiced
(Natural join)

Orders not invoiced (2)

↓

left outer join

O

I

1

1

NULL

2

3

3

4

4

NULL

5

Orders not invoiced (3)

Left outer join

O	I
1	1
2	Null
3	3
4	?
5	Null

where
I. OrderID
is
NULL

Orders not invoiced (4)

O:

1	2	3	4	5	...
---	---	---	---	---	-----

I:

1	3	4	7	9	10	...
---	---	---	---	---	----	-----

Orders and invoices (5)

```
SELECT
  orders.OrderID
FROM
  Sales.Orders AS orders
WHERE
  orders.OrderID NOT IN
  (
    SELECT DISTINCT OrderID
    FROM Sales.Invoices
  )
```

outer
query

1/2/3/4/5

2/5

Subquery (inner
query)

executes first

1/3/4/7/9/10

Existence query (1)

Structure:

→ Outer part

→ subquery ("inner part")

Validates the existence

Existence proof (2)

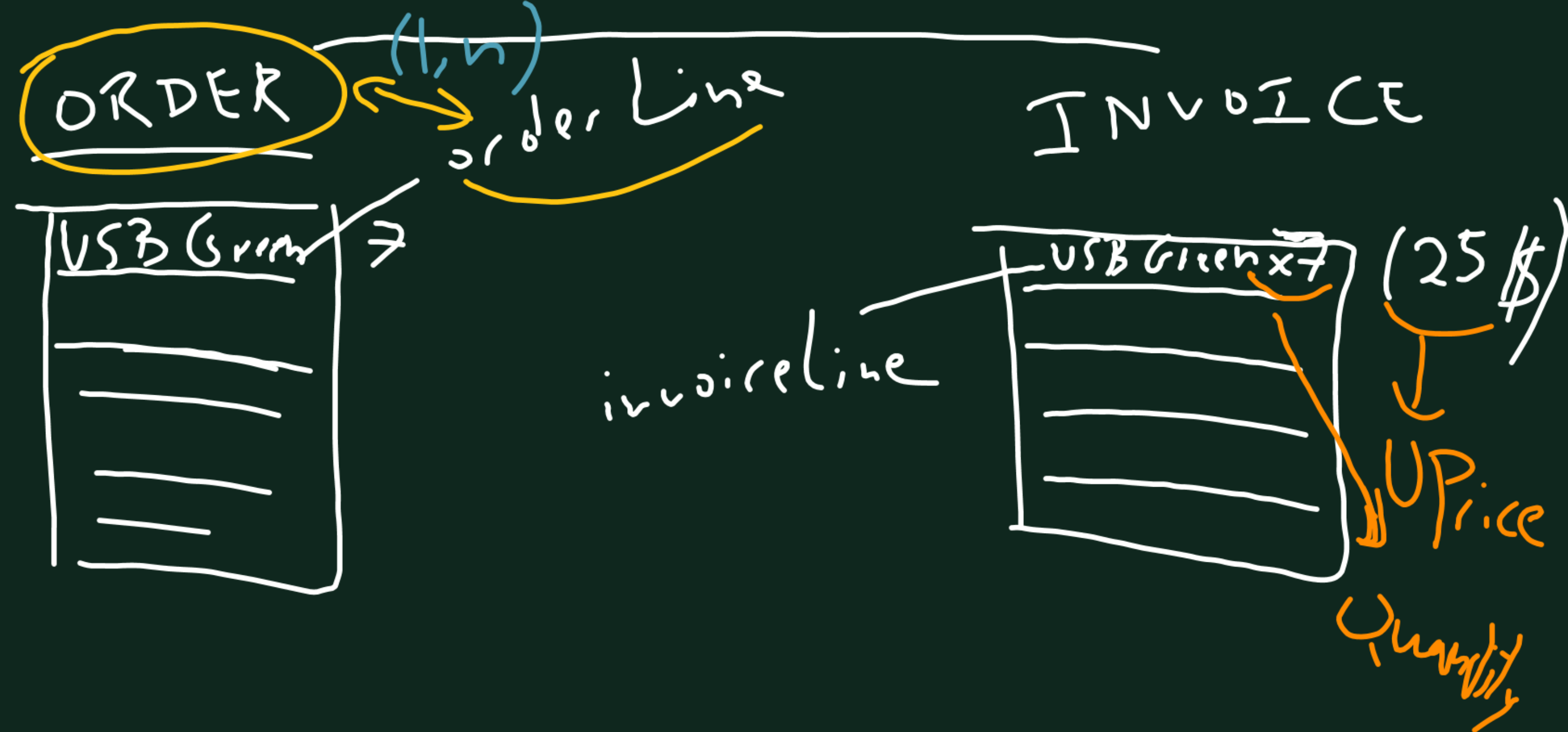
Why we like it:

- Reasoning by 'complement'
(all elements that do not belong to a given set ...)
- Optimal way to solve this kind of problem

Break

Back at 15:20

Orders & Invoices



Duplicate Columns (1)

Attributes (columns)

OLYMPIC

Person	Sport	Height (cm)	Weight (kg)
Sebastian	Running	190	90
Hanna	Swimming	170	170
Hanna	Tennis	170	70

OO

} } #

key

Duplicate Columns (2)

OLYMPIC

PersonID	Sport
1	running
2	Swimming
2	Tennis

ATHLETES

PersonID	Person Name	Height	Weight
1	Sebastian	190	90
2	Hanna	170	70

Date Manipulation

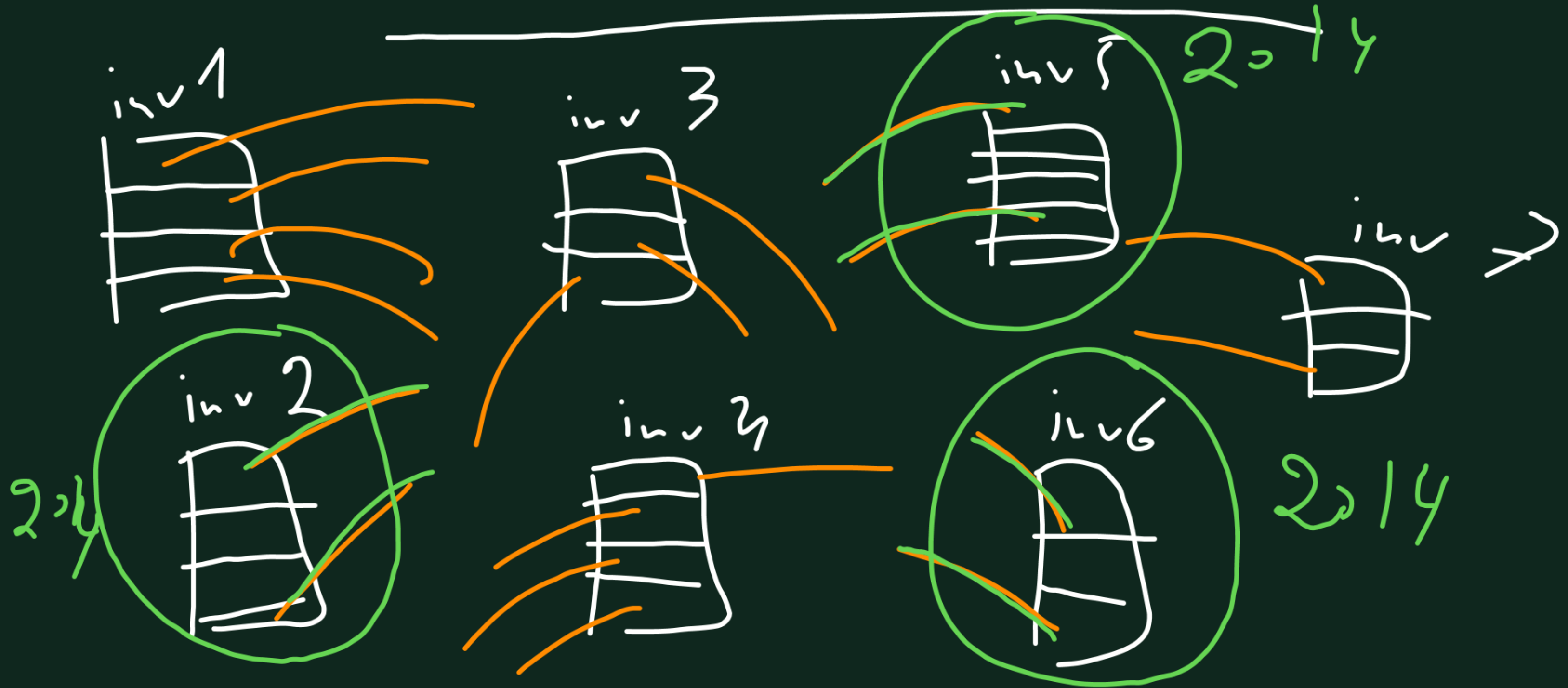
InvoiceDate = '2014-01-01'

① $2014-01-01 \leq \text{InvoiceDate} \leq 2014-12-31$

② $\text{InvoiceDate}(\text{Year}) = 2014$

2014 Profit Query

Invoices & Invoice Lines



↓ Generating a report

2013	2014
105,000	110,000